ADHD Risk Prediction in the Boston Birth Cohort: an application of the Proportional Odds Model and ROC Curve Analysis

by

Ruthe Huang

A thesis submitted to The Johns Hopkins University

in conformity with the requirements for the degree of

Master of Science

Baltimore, Maryland

April 2019

© 2019 by Ruthe Huang

All rights reserved

Abstract

While postnatal lead and mercury levels have been widely studied in relation to neurological disorders in children, the relationship between prenatal lead and mercury exposure and longitudinal childhood outcomes is still unclear. We use the Boston Birth Cohort to illustrate an application of the proportional odds model, an extension of the logistic regression that is often overlooked in time-to-event data analysis. Similar to many events of interest in medical research, diagnosis of attention-deficit/hyperactivity disorder (ADHD) is a binary outcome which occurs at differential time points – if at all – for children. The proportional odds model estimates the effects of covariables on ADHD diagnosis while controlling for the time of diagnosis for each child. We used the proportional odds model to investigate the importance of prenatal lead and mercury exposure biomarkers as well as other prenatal factors in predicting a child's risk of ADHD diagnosis. Additionally, we performed time-dependent ROC curve analysis to compare the prediction performance of eight proportional odds models as well as models created using lasso and a boosted classification tree.

Our proportional odds model does not indicate that prenatal lead and mercury concentrations have substantial effects in predicting odds of ADHD

diagnosis in children. The model does, however, confirm previously known prenatal risk factors for ADHD diagnosis in children, including child sex and race. Furthermore, our ROC curve analysis indicates that prediction tools are not ideal for predicting diagnoses at young ages ($\leq$ 5 years) due to current trends of ADHD diagnoses occurring at ages closer to 6 and 7 years or older. Although our analyses did not find a substantial effect of prenatal lead and mercury on ADHD diagnoses, the mechanisms and exact relationship between these exposure biomarkers – as well as their joint effects with other known risk factors – and neurological outcomes in children still needs more investigation.

Thesis Committee

Primary Readers

Mei-Cheng Wang (Primary Advisor)
Professor
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

Xiaobin Wang
Professor
Department of Population, Family, and Reproductive Health
Johns Hopkins Bloomberg School of Public Health

Acknowledgments

Foremost, I would like to thank my thesis advisor, Dr. Mei-Cheng Wang, for her continuous support of my learning and thesis completion. Her patience, expertise, motivation, and immense dedication to teaching supported my journey through this program and drove me to maximize my learning through my experience. Thank you, Dr. M Wang, for the honor of being your student and advisee.

Additionally, I have immense appreciation for Dr. Xiaobin Wang, whose extensive work in child health inspired my original desire to pursue this topic for my thesis. I give great thanks for the honor of working with Dr. X Wang and for her endless support and patience in helping me to broaden my understanding of prenatal and pediatric health research.

Furthermore, I give great thanks to Dr. Yuelong (Jerry) Ji, who responded to my many questions and data requests with endless patience. I am extremely grateful for his never-ending willingness to help me and the speedy consideration with which he helped me progress in my thesis work.

I would like to thank Dr. Xiumei Hong and Mary Joy Argo for ensuring the logistics of my thesis work moved smoothly and efficiently. I also thank Yuchen Yang for her help and expertise in survival model theory and

applications.

I give great thanks to Dr. Kristin Voegtline for granting me my first research responsibilities and having faith in my abilities. I am truly grateful for the interest she sparked for me in child health and for being such a dedicated, wonderful mentor. I would not be the public health professional and biostatistician I am today without your direction and support.

I am sincerely grateful to my friends and cohortmates – Xinye Li, Andrew Pita, and Emily Scott – for keeping me sane through these two monumental years in the ScM program and for being never-ending providers of support, encouragement, and love.

Finally, and perhaps most importantly, thank you to my parents, Tong Zhang and Shuo Huang, who not only established themselves in a new country but also sacrificed an unthinkable amount to grant their children the gift of education and opportunity.

Table of Contents

Table of Contents vii

List of Tables ix

List of Figures x

1 The Proportional Odds Model 1

1.1 Motivation 1

1.2 Extension of Logistic Regression 1

1.3 Proportional Odds vs. Proportional Hazards 3

2 ROC Methods 5

2.1 Biomarkers 5

2.2 ROC Curves 7

2.3 Time-Dependent ROC Curves 9

2.4 Estimation using the Proportional Odds Model 9

3 Boston Birth Cohort: An application 11

3.1 Introduction 11

3.2 Methods 13
3.3 Results 15
3.3.1 Proportional Odds Model 15
3.3.2 Lasso Model 22
3.3.3 Boosted Classification Tree Model 24
3.3.4 Time-Dependent ROC Curve Analysis 24
3.4 Discussion 28
References 32
Curriculum Vitae 37

List of Tables

3.1 Summary statistics of BBC subset used in analysis 18

3.2 Exponentiated coefficient estimates from proportional odds model 19

3.3 Exponentiated coefficient estimates from lasso model 23

3.4 AUC values for various risk prediction models at four sampling time cutoffs 27

List of Figures

2.1 Receiver Operating Characteristic (ROC) and area under ROC curve (AUC) 8

3.1 ADHD diagnosis prevalence by maternal race of BBC subset used in analysis 16

3.2 Distributions of lead and mercury biomarkers collected from maternal plasma 16

3.3 ADHD diagnosis by prenatal lead concentration, truncated at 10 μg/dL 17

3.4 ADHD diagnosis by prenatal mercury concentration, truncated at 10 μg/L 17

3.5 Overall and sex-stratified baseline odds estimated by full proportional odds model 20

3.6 Relative influence of predictors from boosted classification tree model 24

3.7 ROC curves for various risk prediction models at four sampling time cutoffs 26

Chapter 1

The Proportional Odds Model

1.1 Motivation

Logistic regression is a hugely popular method in biomedical research, as it is a powerful tool when presented with binary outcomes, e.g., presence of disease, death, surgical success, or instances of patient behavior. However, many of these outcomes involve a time-to-event factor that adds valuable information, and thus Cox proportional hazards are often used in attempt to investigate survival times. However, in many cases a proportional hazards model inaccurately describes the problem at hand and the lesser-known proportional odds model is a more appropriate extension of logistic regression. This thesis will use the proportional odds model to predict a binary disease response.

1.2 Extension of Logistic Regression

Suppose that we are interested in a binary outcome, e.g., death, with a p-dimensional vector of covariates $\boldsymbol{Z}$. A logistic regression models the odds

ratio of having the outcome:

$$logit\{P(\boldsymbol{Z})\} = \ln\left\{\frac{P(\boldsymbol{Z})}{1-P(\boldsymbol{Z})}\right\} = \ln\left\{\frac{\text{diseased}}{\text{not diseased}}\right\} = \alpha + \beta'\boldsymbol{Z}$$

The proportional odds model can be regarded as a generalization of the logistic regression model where time-to-event T is involved. Here, we present the proportional odds framework used by Rossini and Tsiatis (Rossini and Tsiatis, 1996). Consider T as being determined by current status data, in which the observation consists only of an examination time and knowledge of whether the event occurred before the exam. This is a common context for medical outcomes since the onset of a condition is recorded through medical diagnosis, which only occurs when a physician examines a patient. Since using current status data is essentially a form of interval censoring, the true time to event is never observed; the observable data consists only of $(X, \Delta, \boldsymbol{Z})$; thus, the exact failure time T is replaced with (X, Δ), where X is the monitoring time and Δ is an indicator of whether the event occurred before the monitoring time, i.e., $\Delta = I\{T \leq X\}$. The proportional odds model is based on the assumption that conditional on $\boldsymbol{Z}$, X is independent of T.

So, the probability that our outcome occurred at or before the monitoring time is $F(\boldsymbol{Z}) = 1 - S(t;\boldsymbol{Z})$, where $S(t)$ is the probability of survival, i.e., the probability of the event not occurring by time t. Under the assumption of proportional odds – i.e., $\phi(t;\boldsymbol{Z}) = \phi_0(t)\exp(\beta'\boldsymbol{Z})$, where $\phi_0(t) = \frac{1-S_0(t)}{S_0(t)}$ is an unspecified baseline odds function – we can apply the proportional odds model:

$$logit\{S(t;\boldsymbol{Z})\} = \ln\left\{\frac{1-S(t;\boldsymbol{Z})}{S(t;\boldsymbol{Z})}\right\} = \ln\left\{\frac{\text{diseased at } t}{\text{not diseased at } t}\right\} = \alpha_0(t) + \boldsymbol{\beta}'\boldsymbol{Z}$$

Note that $\alpha(t)$ is a monotone-increasing function and is interpreted as baseline log odds, or the log odds of the event occurring at time t when the covariates are null, i.e., $\boldsymbol{Z} = 0$. $\boldsymbol{\beta}$ is a p-dimensional regression coefficient, where $\boldsymbol{\beta}'\boldsymbol{Z}$ is the additive change in the log odds of failure with covariates $\boldsymbol{Z}$. The joint distribution of X and $\boldsymbol{Z}$ does not depend on $(\alpha, \boldsymbol{\beta})$ and thus is ancillary. The probability of the event occurring, Δ, given the monitoring time X and covariates $\boldsymbol{Z}$ is

$$P(\Delta = \delta) = F(X|\boldsymbol{Z})^{\delta}(1 - F(X|\boldsymbol{Z}))^{1-\delta}$$

$$= \frac{\exp(\delta(\alpha(X) + \boldsymbol{\beta}'\boldsymbol{Z}))}{1 + \exp((\alpha(X) + \boldsymbol{\beta}'\boldsymbol{Z}))}.$$

Inference for α and $\boldsymbol{\beta}$ is based on the conditional likelihood given X and $\boldsymbol{Z}$:

$$l(\alpha, \boldsymbol{\beta}) = \prod_{i=1}^{n}\left\{\frac{\exp(\delta_i(\alpha(x_i) + \boldsymbol{\beta}'\boldsymbol{z}_i))}{1 + \exp(\alpha(x_i) + \boldsymbol{\beta}'\boldsymbol{z}_i)}\right\}$$

1.3 Proportional Odds vs. Proportional Hazards

The important distinction of why a Cox proportional hazards model is not an appropriate extension for a logistic regression model is that typically for diseases or other binary health states, the outcome is recorded as happening ***by*** time t. For example, a patient is recorded as being diagnosed with ADHD

at time t, but the initial diagnosis and, more importantly, the true onset time is not usually at time t but rather at a point in time before t. In similar fashion, the proportional odds model investigates probability disease status at time t as $1 - S(t_i; \mathbf{Z}_i) = F(t_i; \mathbf{Z}_i)$, or the cumulative probability that disease occurred at any time point before or at t. In contrast, the Cox model describes the outcome of interest as happening ***at*** time t. Using our example outcome of death, a Cox model makes inferences about hazard, or "instantaneous death rate." This can be thought of as the probability of death at a time t given that death has not yet occurred before time t. Formally, hazard is defined as:

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \leq T < t + \delta t | T \geq t)}{\delta t} \right\}$$

Thus we see that instead of regarding the cumulative probability of disease, the proportional hazard model only examines the probability of disease occurring within a small window around t, i.e., essentially occurring at exactly time t.

Chapter 2

ROC Methods

2.1 Biomarkers

A biomarker is a biologically measurable feature that can be used to diagnose or predict a physiologic or pathologic condition. Biomarkers could be collected cross-sectionally or longitudinally over time. for prediction purposes, one could consider using the baseline marker or multiple longitudinal markers. Given a biomarker that is used to diagnose or predict a physiologic or pathologic condition, we consider the following variables:

- D: binary outcome of disease status, determined by gold-standard diagnosis; $D = \begin{cases} 1 & \text{disease present} \\ 0 & \text{disease absent} \end{cases}$
- M: univariate biomarker measurement
  - M_0: biomarker measurement from the control population ($D = 0$)
  - M_1: biomarker measurement from the case population ($D = 1$)

Positive predictive value (PPV) and negative predictive value (NPV) are metrics that directly represent the predictability of biomarkers while taking

into consideration the absolute probability of being diseased or disease-free. Using Moskowitz and Pepe's standardized version of PPV and NPV, we define:

- $PPV(v) = P(D = 1|F_M(M) > v), 0 < v < 1$
- $NPV(v) = P(D = 0|F_M(M) \leq v), 0 < v < 1$

as well as

- True positive rate (sensitivity): $TP(m) = P(M > m|D = 1) = P(M_1 > m)$
- False positive rate (1− specificity): $FP(m) = P(M > m|D = 0) = P(M_0 > m)$

An ideal test has $PPV(m_0) = 1$ and $NPV(m_0) = 1$ for a certain constant m_0, indicating that when using the cutoff value of m_0, the test will always give a positive test result for cases and will always give a negative test result for controls. A useless test will have $PPV(m) = \rho$ and $NPV(m) = 1 - \rho$ for each m, where $\rho = P(D = 1)$, the disease prevalence rate; in this case, M and D are independent. Similarly, an ideal test has $TP(m_0) = 1$ and $FP(m_0) = 0$ for a certain constant m_0, and a useless test will have $TP(m) = FP(m)$.

While true positive rate and false positive rate only focus on marker performance, PPV and NPV take into account the absolute risk of disease. This can easily be seen using Bayes' theorem:

$$PPV(v) = P(D = 1|F_M(M) > v)$$

$$= \frac{P(F_M(M) > v|D = 1)P(D = 1)}{P(F_M(M) > v|D = 1)P(D = 1) + P(F_M(M) > v|D = 0)P(D = 0)}$$

and, similarly,

$$NPV(v) = P(D = 0|F_M(M) \leq v)$$

$$= \frac{P(F_M(M) \leq v|D = 0)P(D = 0)}{P(F_M(M) \leq v|D = 0)P(D = 0) + P(F_M(M) \leq v|D = 1)P(D = 1)}$$

2.2 ROC Curves

Biomarker predictability is typically evaluated by sensitivity and specificity using the Receiver Operating Characteristic (ROC) function (figure 2.1). The ROC function is defined as:

$$ROC(p) = TP[FP^{-1}(p)], p \in (0, 1)$$

The area under the ROC curve (AUC) is:

$$AUC = \int_0^1 ROC(p)dp$$

$ROC(p)$ is monotone increasing in $p \in [0, 1]$. If M is predictive – i.e., $TP(m) > FP(m)$ for all m, the ROC function is above the diagonal line connecting the points $(0, 0)$ and $(1, 1)$. If M is non-predictive – i.e., $TP(m) =$

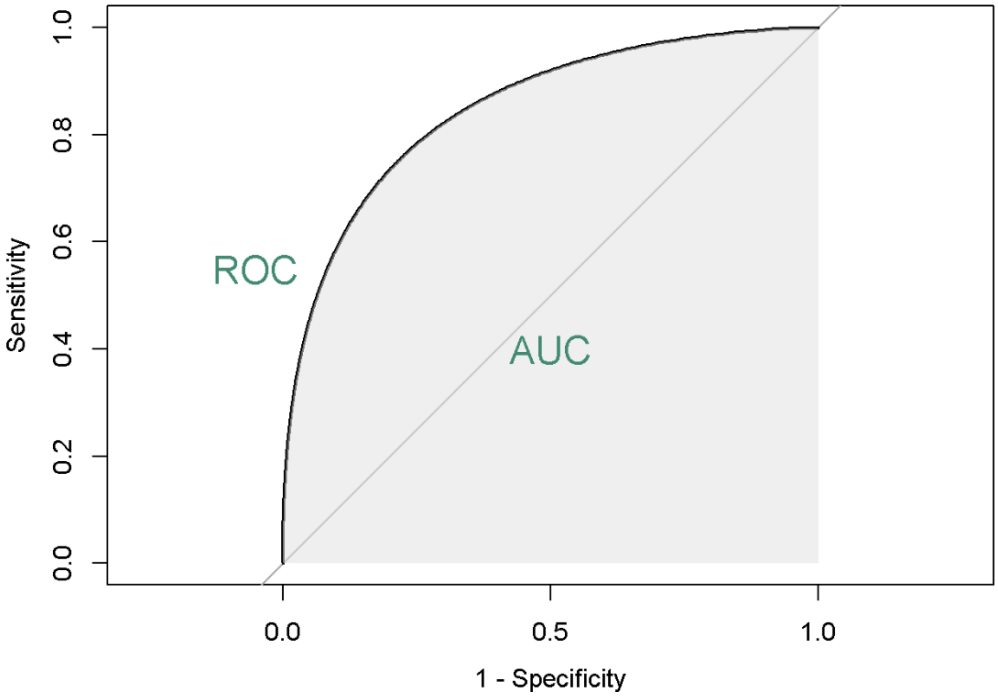


Figure 2.1: Receiver Operating Characteristic (ROC) and area under ROC curve (AUC)

$FP(m)$ – then $ROC(p) = p$ and the ROC function coincides with the diagonal line. In this case, M and D are independent, as discussed previously when $TP(m) = FP(m)$.

Given the logistic regression model, $M = \beta' Z$ maximizes ROC functional values and maximizes the AUC (Alonzo and Pepe, 2002). Similarly, if the proportional odds model holds, then it can be shown that $M \equiv \beta' Z$ yields an optimally combined marker leading to maximized ROC and AUC.

2.3 Time-Dependent ROC Curves

In many cases, the status of disease involves time, extending a binary outcome D_i to time-to-event T_i. The ROC framework can then be extended to the time-dependent ROC, where we use the cutoff $T = t$ as a cutoff to define disease status. Thus the definitions follow:

- Sensitivity: $TP_t^C(m) = P(M > m | T \leq t)$
- 1 - Specificity: $FP_t^D = P(M > m | T > t)$
- $ROC_t^{C/D}(p) = TP_t^C\{[FP_t^D]^{-1}(p)\}$

where D denotes an individual with disease at time t and C denotes an individual without disease and, thus, who is censored at time t.

2.4 Estimation using the Proportional Odds Model

Suppose the proportional odds model holds:

- $\phi(t_i; \boldsymbol{Z}_i) = \frac{1-S(t_i;\boldsymbol{Z}_i)}{S(t_i;\boldsymbol{Z}_i)} = \phi_0(t)\exp(\beta' Z)$
- Define:
  - $M = Z^T\beta$
  - $\tilde{S}(t|Z) = 1 - S(t|Z)$

Then we can derive model-based estimators for TP_t^C and FP_t^D:

$$FP_t^D(m) = P(M > m|T > t) = \frac{\int I(\beta' z > m)S(t|z)dF_Z(z)}{\int S(t|z)dF_Z(z)}$$

$$\approx \frac{\int I(\hat{\beta}' z > m)\{1 + e^{\hat{\alpha}_0(t)+\hat{\beta}' z}\}^{-1}d\hat{F}_Z(z)}{\int\{1 + e^{\hat{\alpha}_0(t)+\hat{\beta}' z}\}^{-1}d\hat{F}_Z(z)} \equiv \hat{FP}_t^D(m)$$

$$TP_t^C = P(M > m|T \leq t) = \frac{\int I(\beta' z > m)\tilde{S}(t|z)dF_Z(z)}{\int \tilde{S}(t|z)dF_Z(z)}$$

$$\approx \frac{\int I(\hat{\beta}' z > m) \times \left\{\frac{e^{\hat{\alpha}_0(t)+\hat{\beta}' z}}{1+e^{\hat{\alpha}_0(t)+\hat{\beta}' z}}\right\}}{\int \left\{\frac{e^{\hat{\alpha}_0(t)+\hat{\beta}' z}}{1+e^{\hat{\alpha}_0(t)+\hat{\beta}' z}}\right\} d\hat{F}_Z(z)} \equiv \hat{TP}_t^D(m)$$

where $\hat{F}_Z$ is the empirical distribution of Z and $\hat{\beta}$ and $\hat{\alpha}_0(t)$ are respectively consistent estimators of β and $\alpha_0(t)$. Note that prospective data is needed for estimation.a

Chapter 3

Boston Birth Cohort: An application

3.1 Introduction

Behavioral problems in children are becoming increasingly prevalent and may lead to adverse outcomes in quality of life, relationships with peers, educational performance, and future professional success (Brauner and Stephens, 2006; Sharma and Couture, 2014). Attention-deficit/hyperactivity disorder (ADHD) is one of the most commonly diagnosed neurodevelopmental disorders among children, affecting 8-12% worldwide (Faraone et al., 2003). However, the etiology of ADHD is still not well understood (Rowland, Lesesne, and Abramowitz, 2002). Recent pediatric research has focused on how the prenatal environment impacts longitudinal childhood outcomes; specifically, increasing evidence points at a relationship between metals and ADHD behaviors. Lead and mercury are among the top three most toxic elements or substances ranked by the US Government Agency for Toxic Substances and Disease Registry (Rice et al., 2014). There exists extensive research on how

lead negatively impacts health in early childhood, as lead interferes with the development of the central nervous system (Canfield et al., 2003). Exposure to high levels of lead may lead to death, coma, and various neurological impairments such as mental retardation, cerebral palsy, and seizures (Hubbs-Tait et al., 2005). Specifically in relation to ADHD, lead has been shown to damage a number of neurotransmitter systems including dopaminergic, glutamatergic, and cholinergic pathways; all three of these pathways are linked to ADHD symptoms (Cory-Slechta, 1995). While there is much evidence regarding the association of postnatal lead levels and ADHD behaviors, there is far less evidence for the relationship between prenatal lead levels and ADHD behaviors (Bellinger et al., 1994; Braun et al., 2006; Needleman et al., 1979).

Mercury can enter the human body through inhaled vapors, bioaccumulation through the food chain, or mercury-contained products such as pigments and preservatives (Rice et al., 2014). Mercury accumulates in nervous tissues throughout the body, but the most detrimental effect of mercury on the nervous system is its interference with energy production, impairing cellular detoxification processes and thus causing the cell to die or live in a chronic state of malnutrition (Rice et al., 2014). Mercury also damages the blood-brain barrier in the central nervous system, facilitating access to the brain by other toxic metals and substances (Rice et al., 2014). Exposure to high concentrations of methyl mercury –a very poisonous form of mercury – has been linked to severe cognitive and motor deficits, seizures, microcephaly, paresthesias, visual loss, and ataxia (Myers et al., 2009; Harada, 1968). The evidence for mercury exposure and ADHD behavior is less consistent than for lead exposure (Kim

et al., 2013). While there is some work showing association between prenatal mercury exposure and ADHD behavior, no conclusion has been strongly established (Boucher et al., 2012; Yoshimasu et al., 2014; Sagiv et al., 2012).

Using a well-established prospective birth cohort and its longitudinal data, we hope to contribute to the understanding of the association between prenatal lead and mercury exposure on childhood ADHD. Additionally, we aim to examine both separate and combined effects of prenatal lead and mercury levels in predicting childhood ADHD. Methodologically, we will apply the proportional odds model and time-dependent ROC curve, instead of the traditional logistic regression model.

3.2 Methods

The data analyzed in this section is from the Boston Birth Cohort. Initiated in 1998, the Boston Birth Cohort (BBC) is a large-scale molecular epidemiological study focusing on environmental factors, genetic variants, epigenomic alterations, and their interactions in adverse pregnancy outcomes. Additionally, a follow-up study of the BBC children to investigate children's health outcomes began in 2002, which prospectively follows mother-infant pairs of the BBC at the Boston Medical Center. The study population consists of predominantly urban, low-income minority and is rich in preterm birth and low birth weight infants. Further details about recruitment and follow-up and research findings have been published previously (Ji et al., 2018; Wang et al., 2014; Wang et al., 2002; Li et al., 2016; Kumar et al., 2011).

We are interested to see if mothers with higher prenatal concentrations

of lead and mercury are more likely to have children with ADHD. Let Y_i denote an indicator random variable that takes the value if the ith child has been diagnosed with ADHD by last follow-up visit on record and 0 otherwise. Let $\boldsymbol{Z}_i$ represent all predictor variables for the ith child, including our main biomarkers of interest – prenatal lead concentration and prenatal mercury concentration – as well as covariables: maternal age, parity, maternal education, maternal race, maternal marital status, maternal stress during pregnancy, alcohol intake during pregnancy, smoking status, delivery method, preterm birth, and child sex. COnvential approaches typically pick a cut-off age, such as age 6, to define binary disease outcomes. In this thesis, we generalize the binary outcome to a time-to-disease outcome, adopt the proportional odds model outlined in Chapter 1, $logit\{S(t_i; \boldsymbol{Z}_i)\} = \alpha_0(t_i) + \boldsymbol{\beta}'\boldsymbol{Z}_i$, and use the gplm package in R version 3.5.3. Furthermore, we created a general model including both female and male children as well as two stratified models by sex to look at any confounding. Additionally, we attempt select machine learning methods – lasso and boosted classification trees – to compare prediction power. We create several models to explore how covariables interact with our biomarkers of interest and then use ROC curve analysis, as outlined in Chapter 2, to gauge which model performed the best for prediction of children's ADHD status.

3.3 Results

3.3.1 Proportional Odds Model

For this study, we used a subset of children from the BBC with available prenatal exposure biomarkers and who were either neurotypical or were diagnosed with ADHD; children with other developmental disorders were excluded from our analysis to isolate our outcome of interest, ADHD. Characteristics of our BBC subsample are shown in Table 3.1. In our sample, boys had higher prevalence of ADHD diagnosis as compared to girls, and white children had higher prevalence as compared to children of other racial groups (figure 3.1). Most mothers had low levels of lead and mercury, with more than 75% of mothers having lead concentration lower than 3.8 μg/dL and 75% of mothers having mercury concentrations lower than 3.7 μg/L (figure 3.2). At first glance, there does not seem to be a clear association between prenatal lead and mercury concentrations and prevalence of ADHD diagnosis (figures 3.3 and 3.4). Lead and mercury concentrations were weakly positively correlated ($\rho = 0.21$).

Table 3.2 shows the odds ratio estimates from our proportional odds model, and figure 3.5 shows the baseline odds of ADHD estimated by the model. By far, the most important factor in determining risk of ADHD in children is sex, with boys having more than 250% higher odds ($OR = 3.605$, 95% CI $[2.579, 5.041]$) of ADHD diagnosis than otherwise similar girls. This sex difference is also seen in the estimated baseline odds (figure 3.5). Other strong risk factors appeared to be maternal marital status and preterm birth. A

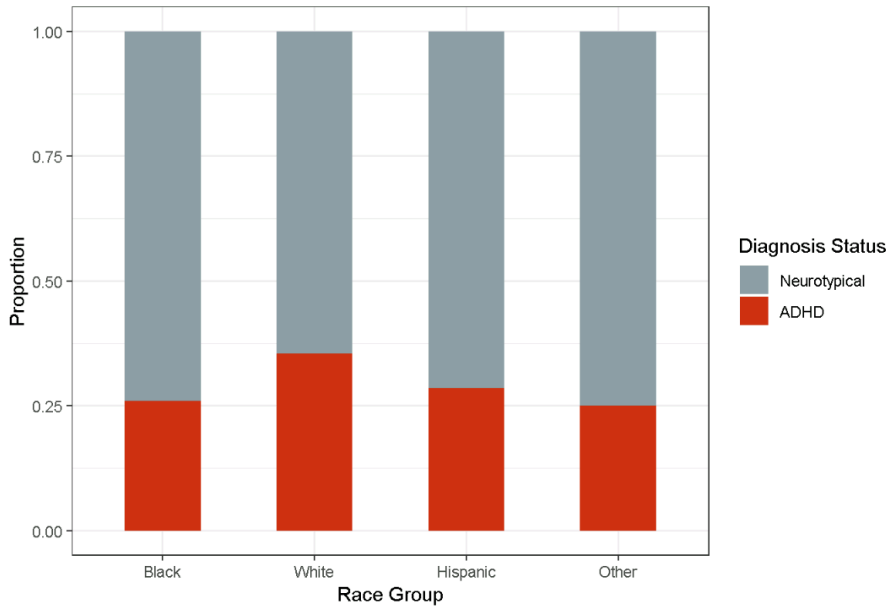


Figure 3.1: ADHD diagnosis prevalence by maternal race of BBC subset used in analysis

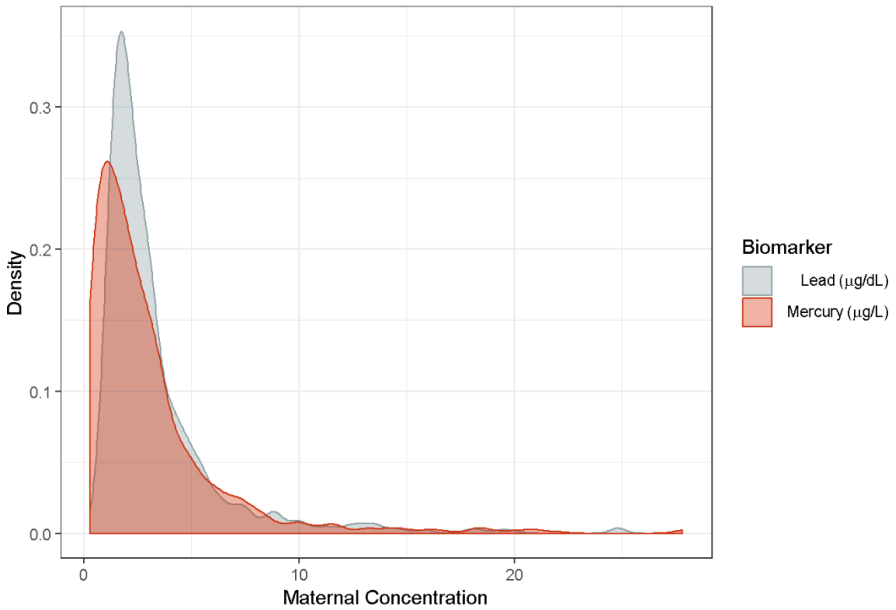


Figure 3.2: Distributions of lead and mercury biomarkers collected from maternal plasma

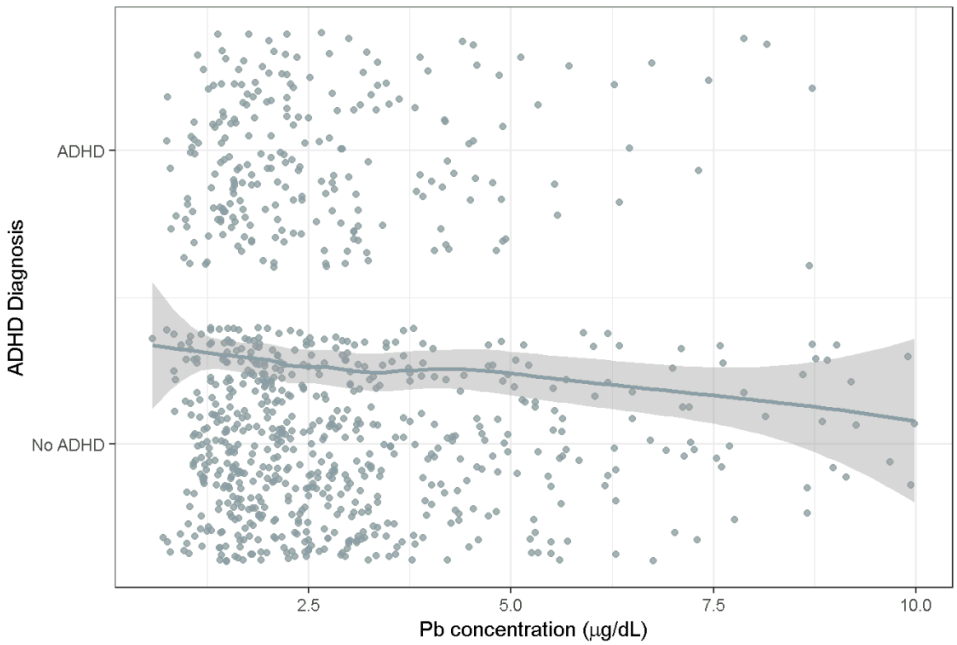


Figure 3.3: ADHD diagnosis by prenatal lead concentration, truncated at 10 μg/dL

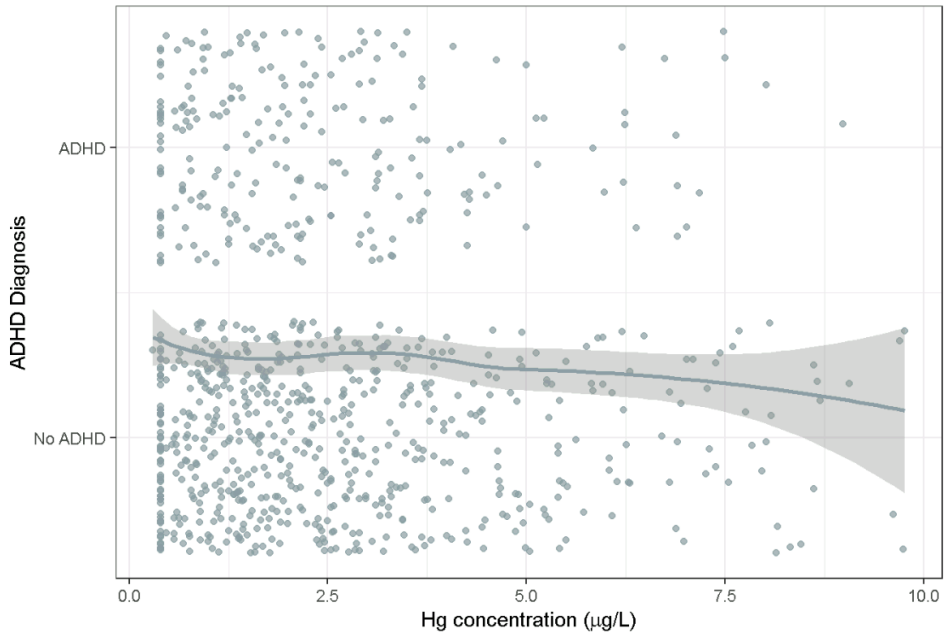


Figure 3.4: ADHD diagnosis by prenatal mercury concentration, truncated at 10 μg/L

Table 3.1: Summary statistics of BBC subset used in analysis

Sample Characteristics	Mean (sd)	n (%)
n		892 (100%)
- Neurotypical		652(73.1%)
- Clinically diagnosed ADHD		240 (26.9%)
Male		455 (51.0%)
Age at last follow-up (years)	8.13 (3.31)	
Age at first ADHD diagnosis (years)	6.05 (2.19)	
Race		
- Black/African-American/Caribbean		598 (67.0%)
- White		48 (5.38%)
- Hispanic		182 (20.4%)
- Other		64 (7.17%)
Preterm Birth		203 (22.8%)
Maternal lead (μg/dL)	3.33 (2.99)	
Maternal mercury (μg/L)	3.06 (3.34)	

child born to a unmarried mother is more than 50% ($OR = 1.516$, 95% CI $[1.03, 2.232]$) higher odds of being diagnosed with ADHD than a child born to a married mother, holding all other variables constant. Our stratified models show, however, that this effect is estimated to be a stronger risk factor in boys ($OR = 1.854$, 95% CI $[1.138, 3.02]$) than for girls ($OR = 1.091$, 95% CI $[0.564, 2.11]$). We also see a sex difference in the effect of preterm birth; a preterm infant is estimated to have more than 60% higher odds ($OR = 1.606$, 95% CI $[1.109, 2.324]$) of ADHD than an otherwise similar term infant, but this effect is estimated to be stronger for boys ($OR = 1.671$, 95% CI $[1.056, 2.646]$) than for girls ($OR = 1.496$, 95% CI $[0.763, 2.934]$).

Other risk factors identified by our model were maternal stress, maternal race, alcohol use during pregnancy, and maternal age; however, the estimated

Table 3.2: Exponentiated coefficient estimates from proportional odds model

	Overall Model		Male-Stratified Model		Female-Stratified Model	
Z	OR Estimate	95% CI	OR Estimate	95% CI	OR Estimate	95% CI
Male	3.605	(2.579, 5.041)	–	–	–	–
Alcohol Use	1.1	(0.621, 1.949)	1.087	(0.525, 2.251)	1.419	(0.547, 3.684)
Vaginal Delivery	0.734	(0.522, 1.032)	0.801	(0.521, 1.231)	0.708	(0.39, 1.285)
Preterm Birth	1.606	(1.109, 2.324)	1.671	(1.056, 2.646)	1.496	(0.763, 2.934)
Stress - Average	1.324	(0.917, 1.911)	1.477	(0.928, 2.351)	1.044	(0.552, 1.975)
Stress - High	1.558	(0.962, 2.524)	1.581	(0.84, 2.975)	1.465	(0.658, 3.261)
Maternal age < 20	1.102	(0.608, 1.998)	0.711	(0.342, 1.478)	2.801	(1.005, 7.804)
Maternal age ≥ 35	1.511	(0.967, 2.36)	1.829	(1.046, 3.197)	1.103	(0.474, 2.571)
Multiparous	0.854	(0.599, 1.218)	0.757	(0.484, 1.185)	0.972	(0.524, 1.803)
Not Married	1.516	(1.03, 2.232)	1.854	(1.138, 3.02)	1.091	(0.564, 2.11)
Above College Education	0.834	(0.581, 1.196)	0.779	(0.49, 1.238)	0.825	(0.445, 1.532)
Smoker - Quitter	1.962	(1.113, 3.458)	1.537	(0.735, 3.214)	3.365	(1.364, 8.3)
Smoker - Continuous	0.889	(0.485, 1.629)	0.481	(0.218, 1.061)	2.178	(0.85, 5.583)
Race - White	1.311	(0.645, 2.667)	0.921	(0.37, 2.293)	2.287	(0.76, 6.884)
Race - Hispanic	0.998	(0.653, 1.525)	0.798	(0.468, 1.361)	1.581	(0.763, 3.279)
Race - Other	1.08	(0.569, 2.05)	1.179	(0.518, 2.683)	1.178	(0.401, 3.458)
Lead (μg/dL)	1.106	(0.959, 1.077)	1.007	(0.939, 1.08)	1.04	(0.937, 1.154)
Mercury (μg/L)	0.941	(0.886, 1)	0.919	(0.849, 0.994)	0.974	(0.882, 1.077)

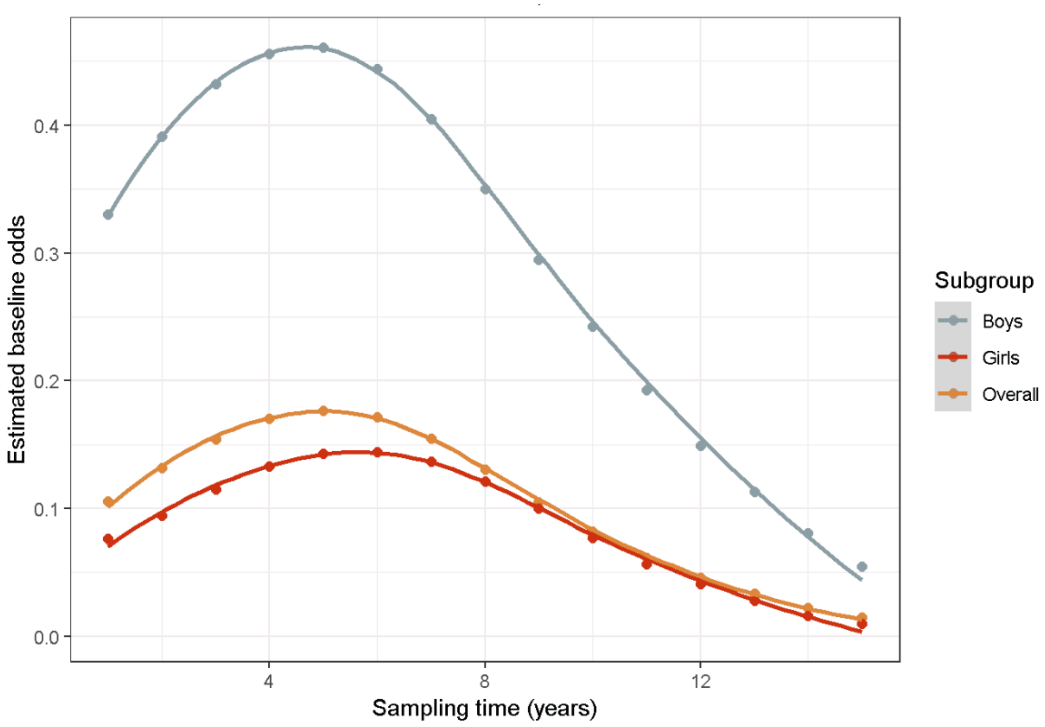


Figure 3.5: Overall and sex-stratified baseline odds estimated by full proportional odds model

effects of these factors were not as strongly supported by the data. Additionally, no substantial sex difference was found among these covariables, with the exception of maternal age. Compared to children of mothers who were not stressed during pregnancy, children of mothers who experienced average stress or were very stressed during pregnancy were respectively estimated to have more than 30% ($OR = 1.324$, 95% CI $[0.917, 1.911]$) and 55% ($OR = 1.558$, 95% CI $[0.962, 2.524]$) higher odds of ADHD diagnosis. Compared to children born to black mothers, those born to white mothers were estimated to have more than 30% ($OR = 1.311$, 95% CI $[0.645, 2.667]$) higher odds; however, children born to Hispanic mothers had similar odds of ADHD diagnosis as children born to black mothers ($OR = 0.998$, 95% CI $[0.653, 1.525]$). Drinking during pregnancy is estimated to have a small magnitude of effect on risk of ADHD, with a child born to a mother who drank having 10% ($OR = 1.100$, 95% CI $[0.6211, 1.949]$) higher odds of ADHD than an otherwise similar child whose mother did not drink during pregnancy. The effect of maternal age on risk of ADHD differed by sex. For girls, being born to a young mother – i.e., a woman less than 20 years of age – increased odds of ADHD by almost 200% ($OR = 2.801$, 95% CI $[1.005, 7, 804]$) as compared to girls born to mothers between the ages of 20 and 35. For boys, being born to an older mother – i.e., a woman at or greater than 35 years of age – increased odds of ADHD by more than 80% ($OR = 1.829$, 95% CI $[1.046, 3.197]$), compared to boys born to mothers between the ages of 20 and 35.

Our model also identified a few protective factors, although none were

considerably supported by the data. Children born by vaginal delivery are estimated to have more than 25% lower odds ($OR = 0.734$, 95% CI $[0.522, 1.032]$) of ADHD than otherwise similar children born by cesarean section. Children who were not their mother's first-born were estimated to have about 15% lower odds ($OR = 0.854$, 95% CI $[0.599, 1.218]$) of ADHD as compared to first-born children. Children whose mothers have above college education were estimated to have almost 17% lower odds ($OR = 0.834$, 95% CI $[0.581, 1.196]$) of ADHD than children of mothers with college or lower-than-college education.

Interestingly, our biomarkers are not associated withx a child's odds of ADHD. A one-μg/dL increase in maternal lead concentration is associated with a mere 1.6% increase ($OR = 1.016$, 95% CI $[0.959, 1.077]$) in odds of ADHD. This increase is estimated to be slightly higher in girls ($OR = 1.040$, 95% CI $[0.937, 1.154]$) than in boys ($OR = 1.007$, 95% CI $[0.939, 1.080]$), but again, the estimated effect is minimal. Similarly, a one-μg/L increase in maternal mercury concentration is associated with an almost 6% decrease ($OR = 0.941$, 95% CI $[0.886, 1.000]$) in odds of ADHD. This decrease was estimated to be larger for boys ($OR = 0.919$, 95% CI $[0.849, 0.994]$) than for girls ($OR = 0.974$, 95% CI $[0.882, 1.077]$), again with little medical significance.

3.3.2 Lasso Model

Our lasso model did not identify lead as an important factor in predicting ADHD diagnosis; additional "unimportant" predictors included low maternal age, multiparity, continuous smoker status, Hispanic race, and other race.

Table 3.3: Exponentiated coefficient estimates from lasso model

Z	OR Estimate
Alcohol Use	1.515
Vaginal Delivery	0.700
Preterm Birth	1.278
Stress - Average	1.081
Stress - High	1.266
Maternal age < 20	.
Maternal age ≥ 35	1.328
Multiparous	.
Not Married	1.337
Above College Education	0.854
Smoker - Quitter	1.143
Smoker - Continuous	.
Race - White	1.540
Race - Hispanic	.
Race - Other	.
Lead (μg/dL)	.
Mercury (μg/L)	1.873
Sampling time (years)	0.873

Similar to our proportional odds model, the lasso model identified mercury as a weak protective factor. Again similar to our proportional odds model, the model also estimated that male sex, preterm birth, white race, alcohol use during pregnancy, maternal stress, maternal age above 35, single mothers, and former smoker status are risk factors and that vaginal delivery and above-college maternal education are protective factors. A table of estimated coefficients from lasso is shown in table 3.3; note that no confidence intervals are provided because there lacks consensus on a statistically valid method of calculating standard errors for lasso predictions (Kyung et al., 2010).

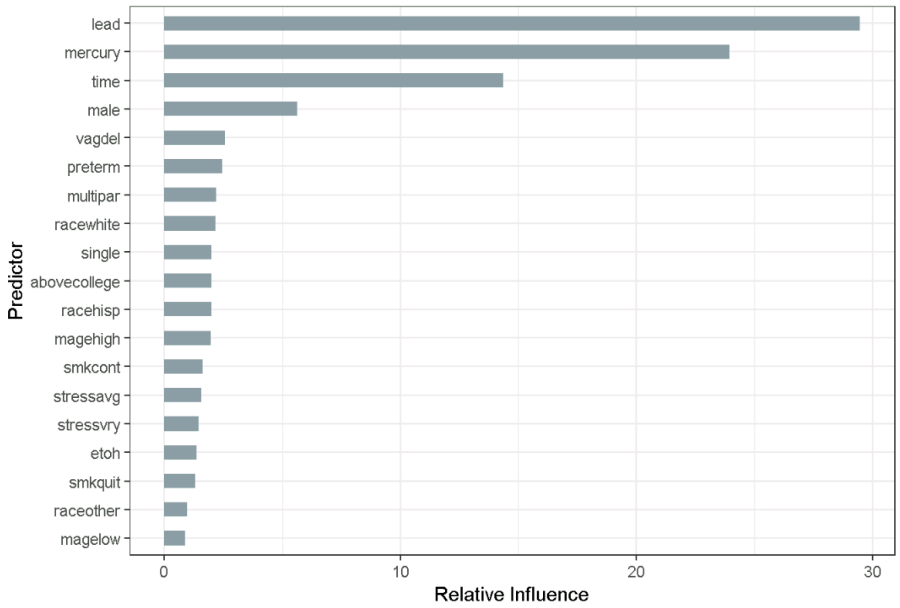


Figure 3.6: Relative influence of predictors from boosted classification tree model

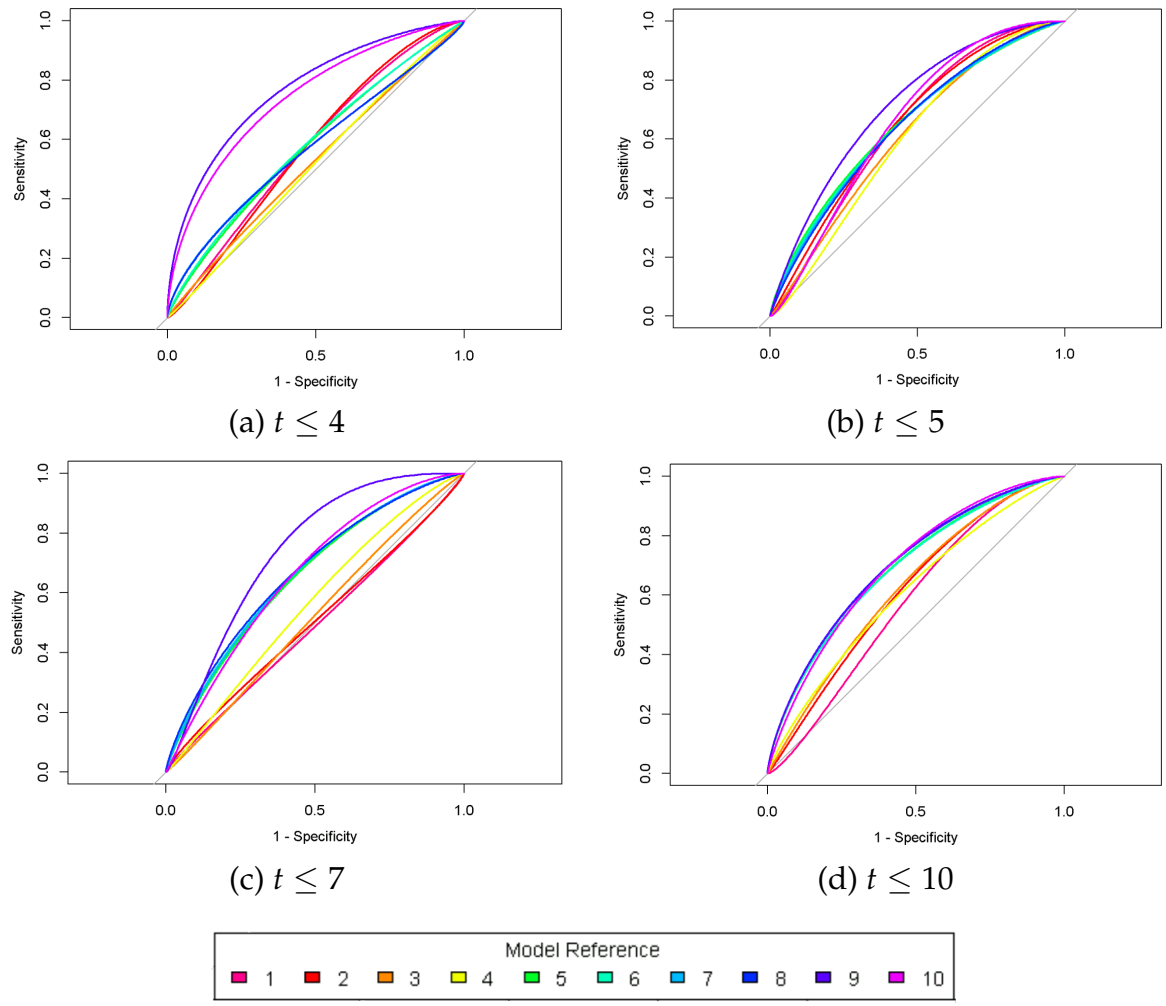
3.3.3 Boosted Classification Tree Model

Unlike our proportional odds and lasso models, our boosted tree model actually identified lead and mercury as having the most influence on predicting ADHD status. Following our biomarkers in relative importance are time to diagnosis, child sex, vaginal delivery, and preterm birth, similar to our other models. A plot of all variables by relative importance is shown in figure 3.6.

3.3.4 Time-Dependent ROC Curve Analysis

We created receiver operating characteristics (ROC) curves for ten different models: eight models used proportional odds modeling, one model used lasso, and one model used a boosted classification tree model. The eight

proportional odds models were nested, with the full model including basic demographics (alcohol use during pregnancy, delivery method, maternal stress during pregnancy, maternal age, parity, maternal education, and maternal race), smoking status, maternal marital status, preterm birth, child sex, mercury, and lead. Details on the specific predictors included in each model are available in table 3.4. Shown in figure 3.7, the ROC curves for each model is plotted for four different sampling time cutoffs: 4, 5, 7, and 10 years. The plot for cutoff time t shows how optimal the given models are at predicting ADHD diagnosis at t years of follow-up. For further detail on model performance, the area-under-the-curve (AUC) values for each model at each sampling time cutoff point is shown in table 3.4. We can see how the lasso and boosted tree models generally outperform any of the proportional odds models; however, this improvement is much more obvious at the 4-year cutoff. By the 10-year cutoff, the difference in performance between the lasso, boosted tree, and full proportional odds models are minimal, with the full proportional odds model actually having a higher AUC value than both the lasso and boosted tree models. Even between the 4- and 5-year cutoffs, we see a striking improvement in model performance for all proportional odds models. Additionally, we can observe that the best-performing proportional odds models – according to AUC values – include lead and/or mercury. However, the improvement in AUC with the addition of lead or mercury (or both) is small in magnitude.



(a) $t \leq 4$ (b) $t \leq 5$

(c) $t \leq 7$ (d) $t \leq 10$

Figure 3.7: ROC curves for various risk prediction models at four sampling time cutoffs

Table 3.4: AUC values for various risk prediction models at four sampling time cutoffs

Reference Number	Model Predictors	Area Under the Curve $t \leq 4$	$t \leq 5$	$t \leq 7$	$t \leq 10$
1	Base demographics + Smoking + Marital status + Preterm birth + Child sex + Lead + Mercury (full model)	0.5836	0.6613	0.6861	0.7152
2	Base demographics + Smoking + Marital status + Preterm birth + Child sex + Lead	0.5819	0.6607	0.6842	0.7146
3	Base demographics + Smoking + Marital status + Preterm birth + Child sex + Mercury	0.6087	0.6679	0.6830	0.7088
4	Base demographics + Smoking + Marital status + Preterm birth + Child sex	0.6062	0.6732	0.6810	0.7091
5	Base demographics + Smoking + Marital status + Preterm birth	0.5125	0.6128	0.5626	0.6067
6	Base demographics + Smoking + Marital status	0.5343	0.6161	0.5063	0.6231
7	Base demographics + Smoking	0.5778	0.6783	0.5135	0.6033
8	Base demographics (alcohol, delivery, stress, age, parity, education, race)	0.5736	0.6613	0.4928	0.5620
9	Lasso Model (same predictors as full model)	0.7926	0.7351	0.7352	0.7088
10	Boosted Tree Model (same predictors as full model)	0.7400	0.6693	0.6596	0.6971

3.4 Discussion

Our finding that ADHD diagnosis is heavily impacted by child sex aligns with previous findings that ADHD is more prevalent in boys (Hermens et al., 2005; Arnett et al., 2015; Andersen and Teicher, 2000). However, ADHD may be overdiagnosed in boys due to gender-biased diagnoses (BruchmÃijller, Margraf, and Schneider, 2012). Additionally, our estimate that white children have higher risk of ADHD diagnosis follows the trend that minority children are less likely to be diagnosed with ADHD than their white counterparts (Morgan et al., 2013). However, diagnoses rates do not indicate differential prevalence of ADHD; previous work shows that lower rates of ADHD diagnosis do not correspond with lower likelihood to display ADHD behaviors among black children and Hispanic children raised in non-English-speaking households (Morgan et al., 2014). Additionally, our findings that preterm birth, alcohol use during pregnancy, smoking status, maternal stress during pregnancy, and maternal age may be risk factors echo findings – albeit of various strengths – in previous work (Sciberras et al., 2017; Banerjee, Middleton, and Faraone, 2007). It is also argued that because prenatal and postnatal risk factors are hard to disentangle, associations between prenatal environmental exposures and longitudinal outcomes may not reflect causation (Thapar and Rutter, 2009).

Our ROC analysis may indicate that prenatal biomarkers and other factors may be more useful in predicting ADHD diagnosis at later years. However, ADHD diagnosis generally occurs after age 6 or 7 and thus the result of our ROC analysis may simply reflect the timing of diagnoses (Barkley, 2003). Collective, these two points indicate that ADHD risk prediction should not be

used at ages younger than 5 or 6 years since these models are not well-trained from lower rates of ADHD diagnoses under age 5. We also find from our ROC analysis that lead and mercury only slightly improve the performance of risk prediction models using proportion odds methods. However, this small improvement may be worthwhile to investigate further; there exist methods to evaluate how seemingly small improvements in AUC can lead to substantial improvement in risk prediction and classification (Pencina, D'Agostino, and Vasan, 2008).

The strength of our work lies in the improved statistical methodology on time-to-ADHD-diagnosis data, expanding on previous published logistic regression results. We also analyze the degree to which biomarkers improve the prediction of ADHD diagnosis at various ages. Proportional odds modeling and ROC curve analysis quantify the low extent of evidence that prenatal lead and mercury exposure biomarkers are associated with longitudinal ADHD diagnosis, adding to the body of research on prenatal indicators of neurological disorders. The limitations of our work remain in model design: we modeled the exposure biomarkers as continuous variables, where the estimated odds ratio of ADHD diagnosis may be small in magnitude due to the small degree of a one-unit change in exposure biomarker levels (1 μg/dL for lead and 1 μg/L for mercury). Future work may involve modeling lead and mercury levels as categorical, such as splitting each exposure biomarker into quartiles or using cutoffs established using biological reasoning. Additionally, further exploration is needed to determine why our boosted classification tree identified lead and mercury as important predictors when no other models

did.

As they currently exist, our models do not indicate that prenatal lead and mercury concentrations have substantial effects in predicting ADHD diagnosis in children. Although the link between these prenatal biomarkers and ADHD in children are not well established in literature, there is still potential of longitudinal neurological outcomes. A prospective study in Yugoslavia found an inverse association between prenatal lead concentration and childhood IQ; furthermore, the authors found that this association was independent of postnatal lead concentration (Wasserman et al., 2000). Similarly, another study in Mexico City found that increased prenatal lead – specifically, elevated lead concentration at 28 weeks of gestation – was associated with decreased child IQ (Lourdes et al., 2006). Analogously, prenatal maternal mercury levels have been inversely associated with children's academic and psychological test scores as well as child IQ (Crump et al., 1998; Axelrad et al., 2007). When considering other neurological outcomes, associations have been found between prenatal lead exposure and schizophrenia, as well as prenatal lead exposure and genomic methylation in cord blood, which may possibly influence the child's long-term epigenetic programming (Mark et al., 2004; Richard et al., 2009). Additionally, prenatal mercury exposure has been linked to increased risk of autism spectrum disorders as well as differential children's expressive language performance (Geier, Kern, and Geier, 2009; Hsi et al., 2014). These results indicate that despite our inability to find a relationship between prenatal lead levels and ADHD diagnosis, prenatal lead exposure may still influence children's neurological development and longitudinal health status.

Furthermore, it is important to note that all of these studies use a variety of outcome measures as well as biological sources for lead or mercury – all factors that may influence the strength of an observed association. These previous findings indicate potential that elevated prenatal lead and/or mercury concentrations may lead to adverse neurological development in children; however, the mechanisms and exact relationship between these biomarkers – as well as their joint effects – and neurological outcomes in children still needs more investigation.

References

Rossini, A. J. and A. A. Tsiatis (1996). "A semiparametric proportional odds regression model for the analysis of current status data". In: *Journal of the American Statistical Association* 91.434, pp. 713–721.

Alonzo, Todd A. and Margaret Sullivan Pepe (2002). "DistributionâĂŘfree ROC analysis using binary regression techniques". In: *Biostatistics* 3.3, pp. 421–432.

Brauner, Cheryl Boydell and Cheryll Bowers Stephens (2006). "Estimating the prevalence of early childhood serious emotional/behavioral disorders: Challenges and recommendations". In: *Public health reports* 121.3, pp. 303–310.

Sharma, Alok and Justin Couture (2014). "A review of the pathophysiology, etiology, and treatment of attention-deficit hyperactivity disorder (ADHD)". In: *Annals of Pharmacotherapy* 48.2, pp. 209–225.

Faraone, Stephen V., Joseph Sergeant, Christopher Gillberg, and Joseph Biederman (2003). "The worldwide prevalence of ADHD: is it an American condition?" In: *World psychiatry* 2.2, p. 104.

Rowland, Andrew S., Catherine A. Lesesne, and Ann J. Abramowitz (2002). "The epidemiology of attentionâĂŘdeficit/hyperactivity disorder (ADHD): a public health view". In: *Mental retardation and developmental disabilities research reviews* 8.3, pp. 162–170.

Rice, Kevin M., Jr Walker Ernest M., Miaozong Wu, Chris Gillette, and Eric R. Blough (2014). "Environmental mercury and its toxic effects". In: *Journal of preventive medicine and public health = Yebang Uihakhoe chi* 47.2, pp. 74–83. DOI: 10.3961/jpmph.2014.47.2.74. URL: https://www.ncbi.nlm.nih.gov/pubmed/24744824https://www.ncbi.nlm.nih.gov/pmc/PMC3988285/.

Canfield, Richard L., Donna A. Kreher, Craig Cornwell, and Charles R. Henderson (2003). "Low-level lead exposure, executive functioning, and learning in early childhood". In: *Child Neuropsychology* 9.1, pp. 35–53.

Hubbs-Tait, Laura, Jack R. Nation, Nancy F. Krebs, and David C. Bellinger (2005). "Neurotoxicants, micronutrients, and social environments: Individual and combined effects on children's development". In: *Psychological Science in the Public Interest* 6.3, pp. 57–121.

Cory-Slechta, D. A. (1995). "Relationships between lead-induced learning impairments and changes in dopaminergic, cholinergic, and glutamatergic

neurotransmitter system functions". In: *Annual Review of Pharmacology and Toxicology* 35.1, pp. 391–415.

Bellinger, David, Alan Leviton, Elizabeth Allred, and Michael Rabinowitz (1994). "Pre-and postnatal lead exposure and behavior problems in school-aged children". In: *Environmental research* 66.1, pp. 12–30.

Braun, Joe M., Robert S. Kahn, Tanya Froehlich, Peggy Auinger, and Bruce P. Lanphear (2006). "Exposures to environmental toxicants and attention deficit hyperactivity disorder in US children". In: *Environmental health perspectives* 114.12, pp. 1904–1909.

Needleman, Herbert L., Charles Gunnoe, Alan Leviton, Robert Reed, Henry Peresie, Cornelius Maher, and Peter Barrett (1979). "Deficits in psychologic and classroom performance of children with elevated dentine lead levels". In: *New England journal of medicine* 300.13, pp. 689–695.

Myers, Gary J., Sally W. Thurston, Alexander T. Pearson, Philip W. Davidson, Christopher Cox, Conrad F. Shamlaye, Elsa Cernichiari, and Thomas W. Clarkson (2009). *Postnatal exposure to methyl mercury from fish consumption: A review and new data from the Seychelles Child Development Study*. DOI: //doi.org/10.1016/j.neuro.2009.01.005. URL: http://www.sciencedirect.com/science/article/pii/S0161813X09000278.

Harada, Y. (1968). "Infantile Minamata disease". In: *Minamata disease*, pp. 73–91.

Kim, Stephani, Monica Arora, Cristina Fernandez, Julio Landero, Joseph Caruso, and Aimin Chen (2013). *Lead, mercury, and cadmium exposure and attention deficit hyperactivity disorder in children*. DOI: //doi.org/10.1016/j.envres.2013.08.008. URL: http://www.sciencedirect.com/science/article/pii/S0013935113001503.

Boucher, Olivier, Sandra W. Jacobson, Pierrich Plusquellec, ÃL'ric Dewailly, Pierre Ayotte, Nadine Forget-Dubois, Joseph L. Jacobson, and Gina Muckle (2012). "Prenatal methylmercury, postnatal lead exposure, and evidence of attention deficit/hyperactivity disorder among Inuit children in Arctic Quebec". In: *Environmental health perspectives* 120.10, pp. 1456–1461.

Yoshimasu, Kouichi, Chikako Kiyohara, Shigeki Takemura, and Kunihiko Nakai (2014). "A meta-analysis of the evidence on the impact of prenatal and early infancy exposures to mercury on autism and attention deficit/hyperactivity disorder in the childhood". In: *Neurotoxicology* 44, pp. 121–131.

Sagiv, Sharon K., Sally W. Thurston, David C. Bellinger, Chitra Amarasiriwardena, and Susan A. Korrick (2012). "Prenatal Exposure to Mercury and

Fish Consumption During Pregnancy and Attention-Deficit/Hyperactivity DisorderâĂŞRelated Behavior in ChildrenPrenatal Exposure to Mercury". In: *Archives of Pediatrics Adolescent Medicine* 166.12, pp. 1123–1131. DOI: 10.1001/archpediatrics.2012.1286. URL: https://doi.org/10.1001/archpediatrics.2012.1286.

Ji, Yuelong, Xiumei Hong, Guoying Wang, Nilanjan Chatterjee, Anne W. Riley, Li-Ching Lee, Pamela J. Surkan, Tami R. Bartell, Barry Zuckerman, and Xiaobin Wang (2018). "A prospective birth cohort study on early childhood lead levels and attention deficit hyperactivity disorder: New insight on sex differences". In: *The Journal of pediatrics* 199, 131. e8.

Wang, Guoying, Sara Divall, Sally Radovick, David Paige, Yi Ning, Zhu Chen, Yuelong Ji, Xiumei Hong, Sheila O. Walker, and Deanna Caruso (2014). "Preterm birth and random plasma insulin levels at birth and in early childhood". In: *Jama* 311.6, pp. 587–596.

Wang, Xiaobin, Barry Zuckerman, Colleen Pearson, Gary Kaufman, Changzhong Chen, Guoying Wang, Tianhua Niu, Paul H. Wise, Howard Bauchner, and Xiping Xu (2002). "Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight". In: *Jama* 287.2, pp. 195–202.

Li, Mengying, M. Daniele Fallin, Anne Riley, Rebecca Landa, Sheila O. Walker, Michael Silverstein, Deanna Caruso, Colleen Pearson, Shannon Kiang, and Jamie Lyn Dahm (2016). "The association of maternal obesity and diabetes with autism and other developmental disabilities". In: *Pediatrics* 137.2, e20152206.

Kumar, Rajesh, Hui-Ju Tsai, Xiumei Hong, Xin Liu, Guoying Wang, Colleen Pearson, Katherin Ortiz, Melanie Fu, Jacqueline A. Pongracic, and Howard Bauchner (2011). "Race, ancestry, and development of food-allergen sensitization in early childhood". In: *Pediatrics* 128.4, e829.

Kyung, Minjung, Jeff Gill, Malay Ghosh, and George Casella (2010). "Penalized regression, standard errors, and Bayesian lassos". In: *Bayesian Analysis* 5.2, pp. 369–411.

Hermens, Daniel F., Michael R. Kohn, Simon D. Clarke, Evian Gordon, and Leanne M. Williams (2005). "Sex differences in adolescent ADHD: findings from concurrent EEG and EDA". In: *Clinical neurophysiology* 116.6, pp. 1455–1463.

Arnett, Anne B., Bruce F. Pennington, Erik G. Willcutt, John C. DeFries, and Richard K. Olson (2015). "Sex differences in ADHD symptom severity". In: *Journal of Child Psychology and Psychiatry* 56.6, pp. 632–639.

Andersen, Susan L. and Martin H. Teicher (2000). "Sex differences in dopamine receptors and their relevance to ADHD". In: *Neuroscience Biobehavioral Reviews* 24.1, pp. 137–141.

BruchmÃijller, Katrin, JÃijrgen Margraf, and Silvia Schneider (2012). "Is ADHD diagnosed in accord with diagnostic criteria? Overdiagnosis and influence of client gender on diagnosis". In: *Journal of consulting and clinical psychology* 80.1, pp. 128–138. DOI: 10.1037/a0026582.

Morgan, Paul L., Jeremy Staff, Marianne M. Hillemeier, George Farkas, and Steven Maczuga (2013). "Racial and ethnic disparities in ADHD diagnosis from kindergarten to eighth grade". In: *Pediatrics* 132.1, p. 85.

Morgan, Paul L., Marianne M. Hillemeier, George Farkas, and Steve Maczuga (2014). "Racial/ethnic disparities in ADHD diagnosis by kindergarten entry". In: *Journal of Child Psychology and Psychiatry* 55.8, pp. 905–913.

Sciberras, Emma, Melissa Mulraney, Desiree Silva, and David Coghill (2017). "Prenatal Risk Factors and the Etiology of ADHDâĂŤReview of Existing Evidence". In: *Current psychiatry reports* 19.1, p. 1. DOI: 10.1007/s11920-017-0753-2. URL: https://doi.org/10.1007/s11920-017-0753-2.

Banerjee, Tania Das, Frank Middleton, and Stephen V. Faraone (2007). "Environmental risk factors for attentionâĂŘdeficit hyperactivity disorder". In: *Acta paediatrica* 96.9, pp. 1269–1274.

Thapar, Anita and Michael Rutter (2009). "Do prenatal risk factors cause psychiatric disorder? Be wary of causal claims". In: *British Journal of Psychiatry* 195.2, pp. 100–101. DOI: 10.1192/bjp.bp.109.062828. URL: https://www.cambridge.org/core/article/do-prenatal-risk-factors-cause-psychiatric-disorder-be-wary-of-causal-claims/ECE94ED2E1A6310CA93CD156508D811C.

Barkley, Russell A. (2003). "Issues in the diagnosis of attention-deficit/hyperactivity disorder in children". In: *Brain and development* 25.2, pp. 77–83.

Pencina, Michael J., Ralph B. D'Agostino, and Ramachandran S. Vasan (2008). "Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond". In: *Statistics in medicine* 27.2, pp. 157–172.

Wasserman, G. A., X. Liu, Dusan Popovac, Pam Factor-Litvak, Jennie Kline, C. Waternaux, N. LoIacono, and J. H. Graziano (2000). *The Yugoslavia Prospective Lead Study: contributions of prenatal and postnatal lead exposure to early intelligence*. DOI: //doi.org/10.1016/S0892-0362(00)00106-9. URL: http://www.sciencedirect.com/science/article/pii/S0892036200001069.

Lourdes, Schnaas, Stephen J. Rothenberg, Flores Maria-Fernanda, Martinez Sandra, Hernandez Carmen, Osorio Erica, Silvia Ruiz Velasco, and Perroni Estela (2006). "Reduced Intellectual Development in Children with Prenatal Lead Exposure". In: *Environmental health perspectives* 114.5, pp. 791–797. DOI: 10.1289/ehp.8552. URL: https://doi.org/10.1289/ehp.8552.

Crump, Kenny S., Tord KjellstrÃűm, Annette M. Shipp, Abraham Silvers, and Alistair Stewart (1998). "Influence of prenatal mercury exposure upon scholastic and psychological test performance: benchmark analysis of a New Zealand cohort". In: *Risk Analysis* 18.6, pp. 701–713.

Axelrad, Daniel A., David C. Bellinger, Louise M. Ryan, and Tracey J. Woodruff (2007). "DoseâĂŞresponse relationship of prenatal mercury exposure and IQ: an integrative analysis of epidemiologic data". In: *Environmental health perspectives* 115.4, pp. 609–615.

Mark, G. A. Opler, S. Brown Alan, Graziano Joseph, Desai Manisha, Zheng Wei, Schaefer Catherine, Factor-Litvak Pamela, and S. Susser Ezra (2004). "Prenatal lead exposure, delta-aminolevulinic acid, and schizophrenia". In: *Environmental health perspectives* 112.5, pp. 548–552. DOI: 10.1289/ehp.6777. URL: https://doi.org/10.1289/ehp.6777.

Richard, Pilsner J., Hu Howard, Ettinger Adrienne, Brisa N. SÃąnchez, Robert O. Wright, Cantonwine David, Lazarus Alicia, Lamadrid-Figueroa HÃľctor, Mercado-GarcÃ■a Adriana, Martha Maria TÃľllez-Rojo, and HernÃąndez-Avila Mauricio (2009). "Influence of Prenatal Lead Exposure on Genomic Methylation of Cord Blood DNA". In: *Environmental health perspectives* 117.9, pp. 1466–1471. DOI: 10.1289/ehp.0800497. URL: https://doi.org/10.1289/ehp.0800497.

Geier, David A., Janet K. Kern, and Mark R. Geier (2009). "A prospective study of prenatal mercury exposure from maternal dental amalgams and autism severity". In: *Acta Neurobiol Exp* 69.2, pp. 189–197.

Hsi, Hsing-Cheng, Chuen-Bin Jiang, Tzu-Hsuan Yang, and Ling-Chu Chien (2014). "The neurological effects of prenatal and postnatal mercury/methylmercury exposure on three-year-old children in Taiwan". In: *Chemosphere* 100, pp. 71–76.